Reviewer A:

It appears this manuscript is for a monograph, but it is quite long, and in many places simply repeating the IPCC SREX or at least paraphrasing it. I think the entire manuscript can be shortened considerably by referring to the SREX where appropriate. I would focus this manuscript more on the existing issues in analyzing extremes in the observed record, including data issues, statistics and the development of indices. Results should be used mainly in the context of how to improve research methods, data etc. One good example is the use of the PDSI for drought. It appears that PDSI is overly sensitive to temperature increases. Also in the observed table by region in the SREX and for drought there were often conflicts between the two main global drought studies (Dai and Sheffield/Wood) resulting in assessments of "inconsistent trends" for many regions where additional literature is absent. Clearly analyzing changes in drought still is an area that needs a lot of work. They state that paleo is beyond their discussion, but for drought especially, paleodrought is a very important subject to include. If even the Dust Bowl was not out of the range of natural variability in the paleo record this should be discussed more.

We thank the reviewer for this comment. We feel that we do present a balanced and nuanced discussion of drought, pointing to the uncertainties that the reviewer mentions. While a discussion of drought in a longer paleo-perspective is beyond the scope that we have chosen for this paper, we do mention this in the introduction and have strengthened the point somewhat by adding referencing a second, recent, paper to support the point. Also, we come back to this in the discussion at the end of the paper, where we now include a sentence that specifically states the importance of the study of paleo-drought for documenting the role of low-frequency climate variability in causing droughts of extended duration. We also cite a review paper on North American paleo drought in this context.

Some minor comments: they should mention the Global Surface Temperature Databank initiative and I would recommend they suggest a similar effort for precipitation, and possibly other variables. Tropical cyclones are discussed and there is a new crowdsourcing effort at NCDC that is having the satellite record reanalyzed for Dvorak intensities which could provide new perspectives on wind changes in TCs. Another question is the role of reanalyses in analyzing extremes. There is some mention mainly in the context of storms but as reanalyses get better there should be a bigger role so this could be a recommendation.

As suggested, we have mentioned the International Surface Temperature Initiative, and have recommended similar efforts for precipitation and other variables. We also mention NCDC's crowd-sourcing initiatives.

We disagree somewhat with the comment on reanalyses given our comments on the quality of the reanalyses in the paper. Our recent experience (e.g., Sillmann et al, 2012a; Donat et al, pers comm) and further publications (e.g. Lorenz and Kunstmann 2012) underscores the concern and indicates that differences in temperature and precipitation extremes between different reanalysis products remain as large as differences between models.

Reviewer B:

We thank the reviewer for his/her constructive suggestions and careful reading of the paper, including a very careful review of the references!

Recommendation: *minor revisions*

General Comments: *This paper reviews the status of extremes research focussing on the historical instrumental record. It highlights the shortcomings of the currently available data for extremes analysis including the extremes databases, process understanding and the currently available statistical techniques for the analysis of extremes. It is well written and draws strongly from the recent chapter on extremes in the IPCC special report on extremes (SREX). In its current form it will provide a useful resource for the climate extremes community although I did think that the authors could have provided a little more information on multi-variate extremes and the approaches for dealing with them, since this is an area where there is an increasing demand for information of how coincident extremes will change.*

We agree that there is increasing demand for information concerning compound, or multi-variable, extremes. The paper is already very long and so we have opted not to add material on this topic. However, we have added a paragraph in the Summary and Recommendations section acknowledging the issue, and pointing to research that contributes to this area.

Below I note a few places where the wording of the manuscript could be clearer or may require minor grammatical revision.

Page 3:

Line 117: The work 'thus' at the beginning of the sentence doesn't seem appropriate for linking the material from the previous sentence, which raised the issue of exposure and vulnerability to what follows in the sentence, which discusses seasonality, duration etc. suggest remove 'thus'.

"Thus" has been replaced with "Also".

Lines 257-262: long sentence. Also the point being made regarding figure 2d is not clear. Could the point regarding the gridding first or later be explained in more detail in relation to the figure for the benefit of the readers who have not read the referenced papers.

The long sentence has been replaced with several shorter sentences. The point about order of operations has been clarified by referring to examples in Figure 2d.

Lines 317-318: change 'by comparison against' to 'in comparison to'

Changed as suggested.

Lines 580-581: reword. Observing systems don't influence cyclone characteristics, rather deficiencies in observing systems lead to errors in the representation of cyclone characteristics over time.

The text has been corrected.

Line 585: '…greenhouse gas forcing from increases…' suggest changing to '… greenhouse gas forcing consisting of increases…'

Wording altered slightly to make it clear that it is radiative forcing that results from greenhouse gas changes.

Line 718: sentence is incomplete

Fixed.

Line 757: 'engineering induced' insert hyphen

Hyphen inserted.

Line 941: Correct spelling of Debernard

Corrected

Line 984: '…evident for the less…' change to '…evident by the less…'

Changed as suggested.

Line 998: '..interpret observations …' -> '…interpreted observations…'

Not changed. "Interpret" is used as a verb here, not as an adjective.

Line 1070: 'This includes….' -> 'These include…'

Corrected.

Line 1096: '… as is clearly evidence from…' add 'd' to evidence.

"Evidence" replaced with "evident".

References:

I could not find IPCC (2011), Kharin et al, 2005, Koster et al, 2004 or van Oldenborgh et al 2012 in the reference list.

Corrections made.

Conversely, the following refs in the ref list did not appear to be cited. Crompton et al, 2011, Johnson and Xie 2010, Katz and Brown 1992; Qian et al 2006.

Corrections made.

Kharin and Zwiers 2005 in ref list is cited Kharin et al, (2005) in text.

Corrected.

Kenyon and Hegerl is cited as 2007 in text but 2008 in ref list

Corrected.

Emanuel is misspelt (Emanual) in Knutson ref

Corrected.

Mesinger et al, 2006 in refs is refered to as Messinger et al, 2006 in text.

Corrected.

Reviewer C:

General Comments

This paper focuses on changes in climate extremes in the recent historical record. It includes consideration of statistical methods for the detection of trends in extremes, as well for the attribution of any such changes. Besides assessing the current state of knowledge on this topic, challenges are identified limiting what can presently be said about the detection and attribution of observed changes in climate extremes.

Given several recent reports (e.g., IPCC) on essentially the same topic, one could naturally question whether yet another review is needed. Nevertheless, the present report is well written and more insightful than many others (in particular, not a dull compendium). While I do have some substantive comments about the emphasis (in my opinion, unduly on indices) and about the style (inconsistent treatment of statistical methods), dealing with these comments should be fairly straightforward.

We thank the reviewer for his/her constructive comments and careful reading of the paper.

Specific Comments

(1) Definition of extreme events

Care is taken to distinguish "moderate extremes" (line 74) from more rare events, with it being argued that "the 90th percentile, for example, would not be considered extreme in a statistical sense" (lines 76-77).

In my view, this distinction is arbitrary and not necessarily justified. Depending on the nature of the data, an analysis based on the 90th percentile might well provide an adequate approximation in terms of extreme value theory.

We have amended the wording slightly to say that the 90^{th} percentile of daily values would not necessarily be considered extreme in a statistical sense. We take the point that in some cases, depending upon the rates of convergence of block maxima or threshold exceedances to their limiting extreme value distributions, values above 90^{th} percentile may, in fact, be well approximated by an extreme value distribution.

(2) Emphasis on indices

Much of the paper is devoted to indices of extreme climate. Perhaps this is not surprising given the present emphasis on this approach in climate change research. Indices have the advantage of convenience, ease in understanding, ability to replicate, etc.

Some serious limitations of extreme indices are duly noted, including not necessarily being designed to facilitate making statistical inferences about trends and providing limited

information (e.g., only applying to a particular threshold). On the other hand, approaches based on extreme value theory are amenable to making statistical inferences about trends and apply, in effect, to multiple thresholds (i.e., all more extreme thresholds than the one used in fitting the statistical model).

This is a good point, which we now discuss in our limitations of the ETCCDI indices.

(3) Treatment of statistical methods

The treatment of statistical methods is uneven:

(i) Sometimes only the results are stated (e.g., "statistically significant linear trends," legend to Fig. 6; also legend to Fig. 8) without any mention of the method used or assumptions made.

The caption for Fig. 6 has been amended to give an indication of how trends were determined and assessed.

Some additional information has been added to the caption for Fig. 8, but the supporting papers do not provide full details of how inferences about regression coefficients were made. The fitted regression model included linear trend and perigean tide components (to account for precession of spring tides), but neither the cited paper, nor earlier papers by these authors, state what assumptions were made concerning the residuals. One assumes that they did not account for serial dependence.

(ii) In other cases, the statistical method is identified (e.g., "Global distribution of linear trends … The trends are calculated using the Mann-Kendall nonparametric trend test," legend to Fig. 7; "Kendall's linear trend estimates," legend to Fig. 5). Presumably, "Mann-Kendall" and "Kendall" refer to the same estimation procedure. In any event, this test does not assume a linear trend; rather, it allows for non-linearity (only assuming a monotone trend). So these two legends need to be corrected.

Captions have been clarified. In both cases, trends were estimated with the Theil-Sen estimator (which is also sometimes known as the Kendall slope estimator). In Figure 5, significance is evaluated using the Mann-Kendall trend test. Figure 7 does not provide an assessment of the significance of the estimated trends, although this is provided in the original paper as a separate panel.

(iii) Occasionally the assumptions on which statistical techniques are based are mentioned (e.g., "spatial and/or temporal independence," line 170), yet usually they are not (at the risk of creating misleading impressions).

Generally the discussion is intended to be illustrative of the scope of work that has been done and the issues that arise, but not to critically appraise the statistical methods used in each case. Those interested in specifics will, no doubt, seek out the cited literature for an authoritative description of the methods used.

The discussion of the spatial modeling of extremes, based on "max stable processes," is surprising given the theoretical nature of the topic (lines 431-432). In my view, considerably more statistical research is needed before this approach will be ready for realistic applications to climate extremes. Further, this approach does not seem necessary if the goal is trend detection, not the statistical modeling of the simultaneous occurrence of extreme climate events across space.

I agree that approach remains difficult to use. I disagree, however, that it is unduly emphasized, and I suspect that the reviewer may not fully understand that detection and attribution is not simply a matter of trend detection in the traditional sense. Not being able to model spatial dependence is currently an impediment to performing "optimal" detection and attribution analyses within the framework of extreme value theory in a manner that corresponds most closely with the techniques used to determine the causes of changes in the mean climate. Note that the paragraph does wind up with a sentence pointing to other approaches that have been used.

(4) Challenges

This section, arguably the most important one, is somewhat disappointing. The identified challenges range from the obvious (i.e., the limitations of the available historical observations, already well documented) to wishful thinking (i.e., "The need for improvements in the reliability of estimators of the attributes of heavy-tailed variables," lines 1022-1023). Much effort has already been devoted to extreme value analysis for heavy-tailed variables, so it is not clear that any substantial improvements are imminent.

Rather than mentioning max stable processes again (lines 1025-1026), other techniques (more promising in the short-term) could be mentioned instead. For instance, the regional analysis approach in hydrology could be extended to estimate a common or smoothly varying trend across space (e.g., Hanel et al., 2009). Such an approach only requires conventional univariate extreme value theory, in combination with a resampling scheme, to detect trends in extremes.

More thought needs to be devoted to this section. Think it is a mistake to simply assume that extreme indices should necessarily continue to be the focus of attention (i.e., "business as usual"). Ideally, this section could be much more visionary, at least considering whether departures from the status quo are warranted.

We thank the reviewer for his/her view, but we beg to differ. I think we have suggested departures from the status quo and have attempted to balance emphasis between (i) the quality and extent of the available data resource, (ii) statistical tools, and importantly (iii) physical process understanding. This section has not focused to any great extent on indices, contrary to the reviewer's last paragraph, while it does point to the need for improvements in statistical tools and in process understanding. Regarding the tools, the proclivity to use block maximum approaches rather than peaks over threshold is often criticized by statisticians, and ameliorating this situation is, I think, relatively low hanging fruit that we point to in this section. Modelling, or otherwise accounting for, the spatial dependence of extremes is important. The objective in this case is not just to obtain better point estimates with tighter uncertainty bounds, although that is

certainly useful, but also to facilitate detection and attribution research. We agree with the reviewer that regional analysis approaches can be useful, and these indeed, are already being used in climate research to various levels of sophistication. We have added two references and some text in the introduction to section 2, including the reference suggested by the reviewer.

Editorial Comments

(1) lines 95-96 ["the ExtREmes toolkit (see http://www.assessment.ucar.edu/toolkit/")]
A more dependable url for the R package "extRemes" is http://cran.r-project.org/web/packages/extRemes/

URL replaced, thank you.

(2) Fig. 1
Strictly speaking, the vertical axis should not be labeled "probability of occurrence." What is actually plotted is the probability density function, which needs to be integrated to obtain a probability (as indicated on the figure).

Yes, we have noticed this as well. A better label on the ordinate would be "relative likelihood of occurrence." Nevertheless, we have opted to reproduce the figure (including its labels) as it has appeared elsewhere.

(3) lines 200-201 ("linked, in one way or another, to societal or ecological impacts")
Disagree with this statement. Not all extreme indices are necessarily linked to impacts (some seem motivated solely by meteorological considerations).

We agree that they are not all equally relevant. We have amended the wording slightly. We now say "linked, to greater or lesser degrees".

(4) line 320 ("where precipitation station density is sufficient to reliably estimate grid box mean")
Not clear what "mean" refers to here. Is it relevant for extreme events?

Thank you. This has been clarified by inserting "daily" ahead of "grid box mean". The concern that is being discussed is about whether there are suitable observational products to allow daily precipitation extremes in models to be evaluated.

(5) lines 492-493 ("it is not clear that they simulate daily intensities that are as heavy-tailed as observed")
Another reference is Fowler et al. (2010).

Thank you, reference added.

(6) line 718 ("Detection of such an anthropogenic influence through the use of tropical")
Text ends abruptly (incomplete sentence).

Missing line reinserted.

(7) lines 1518-1519
*The Sillmann et al. (2011) reference is no longer "in press" (Journal of Climate, **24**, 5899-5913).*

Thank you. The reference has been updated.

References
*Fowler, H.J., D. Cooley, S.R. Sain, and M. Thurston, 2010: Detecting change in UK extreme precipitation using results from the climateprediction.net BBC climate change experiment. Extremes, **13**, 241-267.*
*Hanel, M., T.A. Buishand, and C.A.T. Ferro, 2009: A nonstationary index flood model for precipitation extremes in transient regional climate model simulations. Journal of Geophysical Research, **114**, D15107, doi:10.1029/2009JD011712.*